MSc Project - Reflective Essay

Project Title:	Planetary Lander Using Phasic Policy Gradient Algorithm
Student Name:	Priyanka Prabhath
Student Number:	200599373
Supervisor name:	Dr. Angadh Nanjangud
Programme of Study:	MSc Artificial Intelligence

This project presents a method to solve the rocket landing problem by using the phasic policy gradient algorithm. Phasic policy gradient improves sample efficiency while incorporating the benefits of proximal policy optimization which is the current standard algorithm used for solving different problems in the continuous state and action space domain. Itis seen that with improved sample efficiency, there is a better convergence occurring which leads to the agent learning faster. Space exploration had become important for human kind. One of the crucial problems that need to be solved is landing an exploration object on an extra terrestrial surface. Since terrains of the surfaces are unknown, having the object plan and map out the surroundings is needed.

This project was inspired by the paper [1] which uses the proximal policy optimization algorithm to emulate the planetary powered descent and landing for a space-craft with six degree of Freedom. This document presents the approach taken to develop the project and the practical challenges that were come across during this project. It also projects the contributions the project has towards the future. The final section reflects on my personal development when undertaking this project.

Approach

Preliminary steps were taken to familiarize with the library Pytorch and its various functionalities by using tutorials and various online resources along with setting up the local system with the various tools needed such as an environment and CUDA for deep learning. A solid background needed to be built about my knowledge in reinforcement learning. For this, the book Introduction to Reinforcement learning by Sutton and Barto was of prime importance. Along with the book, online lectures from UCL and Stanford supported the journey of gaining knowledge on the basics of Reinforcement learning. The main tools for building this project were the venv that creates environment and platform to build and Jupyter notebooks with its ease of use and easier debugging of code.

For the environment set up, the OpenAI gym was used. It is a collection of Reinforcement Learning environments' benchmarks ranging from Atari games to simplistic robotic simulations used to test the different RL algorithms making is accessible and convenient to use. LunarLander-v2 environment was used to depict the lander on a planetary surface using the BOX2D simulator and represents a basic 3-DOF system with infinite fuel.

The important part of my project was the study and implementation of the Phasic Policy Gradient algorithm. It is a modification of the current popular Policy Proximal Optimization reinforcement learning algorithms where an extra training phase is introduced. It involved learning the rudimentary algorithms starting from value function optimization techniques and Policy gradient techniques to complex actor critic functions. The system was implemented with different configurations of parameters, activation functions and optimizers.

Practical challenges and limitations

Initial challenges faced during the project was obtaining the right setup environment. Mainly due to versioning differences, there were some tricky parts to rectify. This gave me an insight of the different core modules that run within the system to create different environments for different purposes. Since most of the training had to be done on the local system, the training process was time consuming and dependent on the processor speed.

This project had given me a greater appreciation for mathematics. Solving and deciphering the math needed to understand the problem did take up lot of the day. Through that I had learnt many new mathematical concepts. Many libraries needed to be learnt and understood in order to fully implement the different mathematical concepts that were needed.

The study of the physics of the lander system was beyond the scope of this project.

The main challenge yet to be solved is the right parameters and functions needed. So far the training accumulates rewards that are negative but close to zero. The performance even of the best configuration is way below par than the techniques that have already been implemented.

Critical analysis of relationship between theory and practical work

The reason mankind survives today is due to their inherent need for exploration. With today's advancements, it is possible to explore the regions beyond the stars. This inspired me to work on this project along with my passion for artificial intelligence.

The basic concepts of reinforcement learning i.e. of how the system works with its different components was a good base to design the basic learning architecture. The foundations of the different reinforcement learning methods were crucial in the understanding of the phasic policy gradient algorithm. It is built on the simple base of policy gradient methods modified with the addition of value optimization creating the actor critic technique. Along with that, the concept of trust regions brought about the change in keeping the policy updates within certain limits so that the policy does not deviate from the original track thereby rendering the agent useless.

Contribution and further work

The current implementation of the system is rudimentary and has lots of room for improvement. The phasic policy gradient algorithm shows that it is possible to gain more sample efficiency compared to the Policy proximal optimizer. This would lead to more accuracy in moving the space-craft with more fuel efficiency.

Using much more advance hardware, the training of the algorithm would be much more effective. There is a need to improve the structure of the neural network to further optimize the rewards obtained. Currently only the TanH and ReLU activation functions are used. Using better activation functions like the different variations of the ReLU function can be considered along with modifying the structure of the network.

Since this implementation only considers limited environment, the agent training can be modified for a much more realistic environment where the lander can move in all diretions of the plane with a 6-DOF system as well as limited fuel.

Personal development

This project has helped me gain insights to a lot of different aspects of the different scenarios Reinforcement learning can be used. It has helped solve many problems and automate many tasks driving the advancements to the future. I had gained many technical skills on working with the project. It has made me aware of the simple hardware components that are used to drive complex problem solving. Pytorch has helped me gain insight into the working of the neural network structure and the different structures that can be created.

I have also learnt to be disciplined and precise about the work I do. It has helped me gain organisational skills as well as the technical skills. Also, it has helped me meet and network with people in the field which has opened many doors into this field. I have gained a bit of insight on the industrial side of the field.

Overall this project has helped me grow from the basic knowledge to implementing complex algorithms along with gaining analytical skills which helps in experimentation and analysis of the different scenarios possible.